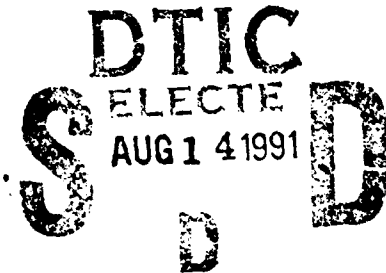


AD-A239 457



2

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022



**Practical Aspects
of an Interior-Point Method
for Convex Programming**

by
Florian Jarre and Michael A. Saunders

TECHNICAL REPORT SOL 91-9

July 1991

91-07752

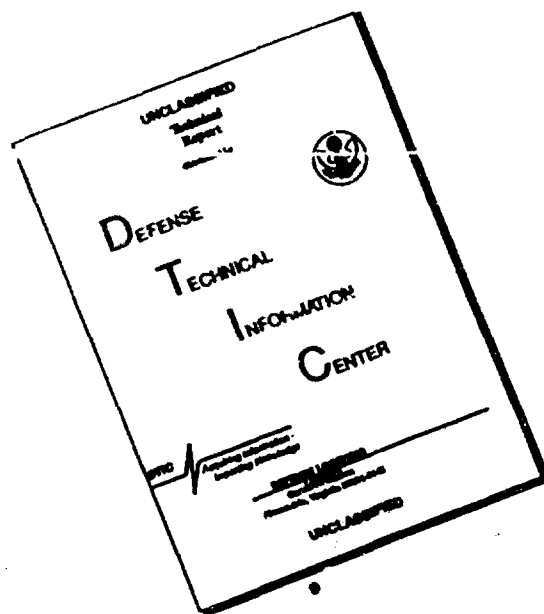


This work was supported by a research grant from the Deutsche Forschungsgemeinschaft, and by the National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do NOT necessarily reflect the views of the above sponsors.

Reproduction in whole or in part is permitted for any purposes of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST
QUALITY AVAILABLE. THE COPY
FURNISHED TO DTIC CONTAINED
A SIGNIFICANT NUMBER OF
PAGES WHICH DO NOT
REPRODUCE LEGIBLY.

PRACTICAL ASPECTS OF AN INTERIOR-POINT METHOD FOR CONVEX PROGRAMMING

Florian Jarre* and Michael A. Saunders

TECHNICAL REPORT SOL 91-9[†]

July 1991

Accession For	
NTIS	CRA&I
DTIC	TAB
Unannounced	
Justification	
By	
Distribution/	
Availability Code	
Dist	Availability or Special
A-1	

Abstract

We present an algorithm for solving convex programs with nonlinear constraints. The algorithm works in the primal space only and uses a predictor-corrector strategy to follow a smooth path that leads to an optimal solution. The algorithm simultaneously iterates towards feasibility and optimality. The matrices occurring in the algorithm can be kept sparse if the nonlinear functions are separable or depend on few variables only. Some promising numerical examples obtained from a preliminary implementation are included.

Key words: convex program, interior-point method, implementation, sparsity

1. INTRODUCTION

Following Karmarkar's proof of polynomial-time complexity of an interior-point method for solving linear programs [10], efficient numerical implementations for the solution of linear programs by related interior-point methods have been presented; e.g. [14, 15, 18] and many others. The applicability of such interior-point methods to *nonlinear convex* programs—like problem (CP) below—was soon recognized by Sonnevend in [28], and a detailed complexity analysis of these applications was presented in [5, 6, 16, 23, 24], showing that for certain classes of nonlinear constraints essentially the same speed of convergence can be expected as for a linear program.

The conversion of these theoretical results into numerical algorithms has been very slow so far. In this paper, we intend to provide some encouragement (supported by numerical experiments) that interior-point methods are in practice—not just in theory—an efficient tool for solving certain classes of convex (and possibly non-convex) problems.

[†]This work was supported by a research grant from the Deutsche Forschungsgemeinschaft, and by the U.S. National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

*On leave from Institut für Angewandte Mathematik, University of Würzburg, 8700 Würzburg, (West) Germany.

1.1. The Problem

The problem under study is to find

$$\min\{f_0(x) \mid x \in P\}, \quad (CP)$$

and a corresponding optimal solution x^* , where the feasible domain¹ P is given by

$$P := \{x \in \mathbb{R}^n \mid f_i(x) \leq 0, \ 1 \leq i \leq m\}, \quad (1.1)$$

and the $f_i : S \rightarrow \mathbb{R}$ ($0 \leq i \leq m$) are continuous convex functions, i.e. $f_i \in C(S)$. We assume that the functions f_i are defined on a common closed set $S \supset P$ and that the f_i are twice continuously differentiable in the interior S° of S . The first and second derivatives are assumed to be known and are denoted by column vectors and square matrices:

$$g_i(x) = \nabla f_i(x), \quad H_i(x) = \nabla^2 f_i(x).$$

The algorithm is designed to work properly even if P is empty or unbounded. The only assumption is that we are given a starting point x^0 in S° . The point x^0 may be infeasible for (CP) .

2. A SIMPLE BARRIER METHOD

2.1. Barrier Methods in General

The principle of a barrier method for approximating the solution of problem (CP) is based on the ideas in [1, 2] and can be outlined as follows.

For $\mu = \mu_0, \mu_1, \mu_2, \dots$ (where $1 = \mu_0 > \mu_1 > \mu_2 \dots$ and $\mu_k \rightarrow 0$), find

$$x(\mu) := \arg \min f_0(x) + \mu\phi(x),$$

i.e. minimize the true objective function f_0 perturbed by $\mu\phi$, where $\phi(x)$ is a smooth convex "barrier function" for the set P (tending to infinity as x approaches the boundary of P and finite in P°).

We will choose the function $\phi(x) = -\sum_{i=1}^m \log(-f_i(x))$. For this definition of ϕ it is straightforward to see that ϕ is smooth and convex if the f_i are so.

It is well known (see e.g. [1]) that the minimizers $x(\mu)$ are unique if, for example, P is bounded, ϕ is strictly convex and f_0 is convex. Further, for any $\mu > 0$ the barrier term $\mu\phi(x)$ ensures that $x(\mu)$ is feasible for (CP) . Fiacco and McCormick showed under weak assumptions that the minimizers $x(\mu)$ converge to an optimal solution of problem (CP) as the perturbation $\mu\phi(x)$ of the objective function is "phased out", i.e. as $\mu \rightarrow 0$.

¹Here we assume a simple (but general) form of P . It is straightforward to include linear constraints and upper and lower bounds on the variables in an efficient way.

2.2. What Makes Them Work

In what follows we try to motivate rather informally what makes a barrier method work.

1. An immediate advantage of a barrier method is that a constrained optimization problem is solved via smooth unconstrained problems. This idea was early recognized in [1, 2].
2. A second important point is that under certain conditions [6, 23] all subproblems are of the same "difficulty" or the same structure; that is, no matter how small μ_k is, the domain of convergence of Newton's method for finding $x(\mu_k)$ is always a "fixed percentage" of the previous level set $\{x \mid x \in P, f_0(x) \leq f_0(x(\mu_{k-1}))\}$, and the "percentage" is independent of the data of the problem. This fact was discovered only recently, and it depends on the choice of the barrier function and the constraint functions. For a detailed analysis we refer to [7, 24].

Note that the minimizers $x(\mu)$ converge to an optimal solution of (CP) , which usually lies at the boundary of P . Since the barrier function goes to infinity as its argument approaches the boundary, one might expect that the subproblems become increasingly harder to solve as μ tends to zero. The above statement is rather surprising in that it guarantees that—from the point of view of the size of the domain of convergence of Newton's method—all subproblems are equally hard to solve (at least in the absence of rounding errors).

3. A third nice property is that (under weak conditions) $x(\mu)$ is a smooth curve in μ (see [1]), and the tangent to $x(\mu)$ at $\mu = \mu_k$ (pointing to the "next point" $x(\mu_{k+1})$) is easily computable (see Section 4.1).
4. Finally it is also important that under weak conditions, one can show that the "distance" $|f_0(x(\mu)) - f_0(x^*)|$ is $O(\mu)$, making the size of μ a reliable stopping criterion.

Below we state a conceptual interior-point algorithm. The algorithm is guaranteed to converge globally with a linear rate of convergence that is independent of the problem data if the second derivatives of the constraint functions satisfy a relative Lipschitz condition as defined in [6], or if the logarithmic barrier function satisfies a so-called self-concordance relation defined in [24]. These conditions guarantee that point 2 above holds true. The complexity analyses based on these conditions (as well as the conditions themselves) are quite involved. Here we only mention that a fairly general class of convex constraints (in particular linear or convex quadratic functions) satisfy these conditions. Our numerical experiments suggest that the algorithm also works for larger classes of problems for which a theoretical proof of convergence has not been shown yet. In this paper we do not assume either of the two conditions given in [7] or [24], and will not give any complexity results.

2.3. Outline of a General Barrier Method

A general outline for a barrier method can be stated as follows.

Assume $\mu_0 = 1$ and $x(\mu_0)$ are given. Set $k = 0$.

Do until convergence

- Compute the tangent $x'(\mu_k)$.
- Select $\mu_{k+1} < \mu_k$.
- Determine a prediction for the next iterate by

$$\hat{x}(\mu_{k+1}) := x(\mu_k) + (\mu_{k+1} - \mu_k)x'(\mu_k).$$

- Find $x(\mu_{k+1})$ by Newton's method starting from $\hat{x}(\mu_{k+1})$.
- Set $k = k + 1$.

End

In this outline we suppressed a number of details that we mention briefly, postponing their detailed discussion to Section 3.

1. The points $x(\mu)$ need not be computed exactly.
2. Newton's method and the extrapolation must be secured by a linesearch (since the function ϕ that defines $x(\mu)$ is not defined outside P).
3. We must find an initial point $x(\mu_0)$ minimizing $f_0(x) + \mu_0\phi(x)$.

2.4. Note on Primal-Dual Methods

The method outlined above works in the primal space only. We briefly mention the relationship to primal-dual methods.

It is straightforward to convert the KKT-conditions of a convex optimization problem into a nonlinear complementarity problem that involves primal and dual variables. The functions defining the nonlinear complementarity problem are monotone (they are the gradients of convex functions) and interior-point algorithms for solving nonlinear complementarity problems with monotone functions have been proposed in [12, 13]. Implementations of such primal-dual methods proved to be very effective when applied to linear programs [14, 15], and it may be expected that the same also holds for nonlinear problems. However, the search directions generated by primal and primal-dual methods coincide at points on the path of analytic centers, and the worst-case complexity for primal and primal-dual algorithms is the same. Moreover, the strong theoretical results proved in [7, 24] about the convergence of primal methods for solving convex programs have not been proved (yet) for primal-dual methods.

2.5. Important Details

Our method has much in common with the traditional barrier methods suggested by [2, 1] and implemented in SUMT in 1964. It is natural to ask why these methods did not retain their initial popularity. We try to give a partial answer by pointing out some new developments.

- It is important which barrier function is chosen. For large classes of problems the logarithmic barrier function allows rigorous proofs of convergence that cannot be given for other barriers like $1/f_i(x)$. An implementation that takes advantage of the strong theoretical results—in particular those shown in [24, 7]—may yield a robust solver for some classes of (almost) convex problems.
- As pointed out in [20, 30], the Hessians of the barrier functions become increasingly ill-conditioned² as the iterates approach an optimal solution x^* if there are less than n linearly independent constraints active at x^* . This difficulty also occurs when solving degenerate linear programs by interior-point methods. The large number of numerical experiments carried out to date suggests, however, that with a careful choice of algebra for solving the linear equations this difficulty can be overcome. We may hope that this is also the case when solving nonlinear problems. In addition, the fact that computers today use a much higher arithmetic precision than 25 years ago makes current codes less sensitive to ill-conditioning. Finally, as pointed out by [3], a concept used in early barrier methods of enforcing equality constraints by a quadratic penalty function (rather than linearizing them at each step and using projections) introduced further numerical instability.
- Many implementations of interior-point methods for the conceptually simpler problem of solving linear programs have been tested in the recent past. These implementations documented the great importance of good sparse-matrix techniques. Without the latter, interior-point methods for large linear programs are completely unattractive, and the same may be true for nonlinear problems. It may be anticipated, however, that interior-point methods applied to certain classes of “inherently sparse” (e.g. separable) problems with cheap first and second derivatives will be able to exploit the additional structure and yield fast special-purpose solvers.

3. A MODIFIED BARRIER METHOD FOR (CP)

3.1. Shifted Constraints $f_i(x, \mu)$

A given initial point $x^0 \in S^o$ might not be feasible for (CP). In order to define a barrier function for x^0 we “enlarge” the feasible set P by subtracting certain

²This can also be seen from the inner and outer ellipsoidal approximations of the level sets by ellipsoids given by the Hessian of the barrier function and centered at the points $x(\mu)$. If there are less than n active constraints, the level sets become “flat” and the ellipsoids approximating them hence become singular in the limit; see [7].

nonnegative quantities β_i from f_i such that $f_i(x^0) - \beta_i < 0$, i.e. such that x^0 is in the "enlarged" feasible domain $\{x \mid f_i(x) - \beta_i \leq 0\}$.

More precisely, let $t = \max_{1 \leq i \leq m} \{1, f_i(x^0)\}$ and $\beta_i = \max\{f_i(x^0) + t, 0\}$ for $1 \leq i \leq m$, and define³

$$f_i(x, \mu) = f_i(x) - \mu\beta_i. \quad (3.1)$$

This implies that

$$f_i(x^0, 1) \leq -t, \quad 1 \leq i \leq m.$$

The above computation of β_i is not invariant under multiplication of f_i with a positive constant⁴. Without loss of generality we therefore assume that all functions f_i satisfy

$$\|g_i(x^0)\|_2 + \frac{1}{\sqrt{n}} \|H_i(x^0)\|_F = 1, \quad 0 \leq i \leq m, \quad (3.2)$$

where the Frobenius norm $\|A\|_F$ of a matrix A is given by $\|A\|_F = (\sum A_{ij}^2)^{1/2}$ and is easy to compute. (To arrange this we simply multiply $f_i(x)$ by a suitable scalar before we start the algorithm.) Note that $\nabla_x f_i(x, \mu) = g_i(x)$.

3.2. Shifted Sets P_μ

For $\mu \in [0, 1]$ we consider feasible domains P_μ defined by

$$P_\mu := \{x \mid f_i(x, \mu) \leq 0, \quad 1 \leq i \leq m\}. \quad (3.3)$$

Note that $P_0 = P$ and that x^0 is in the interior of P_1 (i.e. $x^0 \in P_1^\circ$). The following lemma relates the feasible domains P_μ to P .

Lemma 1

1. $P \subset P_{\mu_1} \subset P_{\mu_2}$ for $0 \leq \mu_1 \leq \mu_2 \leq 1$.
2. $P = \cap_{\mu > 0} P_\mu$.
3. If P is not empty, the interior P_μ° is not empty for all $\mu \in (0, 1]$.
4. If P is empty, there is a δ in $[0, 1)$ such that P_μ is empty for all $\mu \in [0, \delta)$ and the interior P_μ° is not empty for all $\mu \in (\delta, 1]$.

Proof: See Appendix B. ■

The algorithm below follows a path of points $x(\mu) \in P_\mu$ from $\mu = 1$ to $\mu = 0^+$. In contrast to the simple outline given earlier, the feasible sets P_μ for the subproblems

³The approach presented here may need to be modified in the following case. A certain function f_i may be convex (or have a self-concordant barrier function) in the domain $\{x \mid f_i(x) \leq 0\}$ but not in $\{x \mid f_i(x) \leq \beta_i\}$ if $\beta_i > 0$. For a more precise statement we refer to Appendix A.

⁴Nor is it affine invariant. However it guarantees that the initial point x^0 is at least $t \geq 1$ away from each constraint $f_i(x) - \beta_i \leq 0$, which is sufficient for our purposes. (Finite-precision arithmetic is not affine invariant either.) An affine invariant computation of β is given by the solution of the following problem: $\min\{\|\beta\|_2 \mid \beta \geq 0, f_i(x^0) - \beta_i < 0, \sum g_i(x^0)/(f_i(x^0) - \beta_i) = 0\}$.

of finding $x(\mu)$ do not remain constant in our method below. For the sets P_μ we define a barrier function ϕ of the free variables x and the parameter μ by⁵

$$\phi(x, \mu) = - \sum_{i=1}^m \log(-f_i(x, \mu)).$$

3.3. The Perturbed Center

Before stating the subproblems that define the points $x(\mu)$, we digress briefly for a technical but convenient detail.

First note that the properties of the barrier function ϕ , in particular its second derivative and convexity, finiteness in P_μ and the limit as x approaches the boundary of P_μ , are invariant under linear perturbations of ϕ .

Hence, the set $\{\tilde{\phi} \mid \tilde{\phi}(x, \mu) := \phi(x, \mu) + w^T x, w \in \mathbb{R}^n \text{ fixed}\}$ of linear perturbations of ϕ forms a family of barrier functions for P_μ . Each barrier function defines a smooth path of minimizers $\arg \min \{f_0(x) + \mu \tilde{\phi}(x, \mu)\}$ leading from some point in P_1^0 to an optimal solution of (CP) . Also, for any $\mu^0 > 0$ and any $x^0 \in P_{\mu^0}^o$ there is a unique w such that the path starts at x^0 when μ runs from μ^0 to 0. Therefore, the functions $\tilde{\phi}$ define a vector field that flows to the optimal set, a fact that will be used extensively later on and is well described for the case of linear constraints in [17]. The minimizers of the perturbed barrier functions will be called *perturbed centers*, and the paths of the vector field will be referred to as *perturbed center paths*.

The "perturbed center" without perturbation (i.e. with $w = 0$) is the *analytic center* defined by Sonnevend in [28]. It exists if, for example, the set of optimal solutions is bounded and the Hessians of the constraints satisfy a relative Lipschitz condition [6]. For the analytic center a number of nice properties can be shown; in particular,

- If the shifts β_i are zero, one can show that there is a two-sided ellipsoidal approximation of the level set $\{x \in P \mid f_0(x) \leq \lambda(\mu)\}$ centered at the point $x(\mu)$, such that the ratio of the inner ellipsoid to the outer ellipsoid is of order m (the number of constraints) independent of the data of the problem [29, 7]. Here, $\lambda(\mu) > f_0(x(\mu))$ is a suitable number whose derivation is explained in [29].
- Again if $\beta_i = 0$, the numbers $\mu/f_i(x(\mu))$ define dual feasible variables that can be used in a test for optimality.

Both properties hold in a somewhat weaker form if the perturbation w is small [6].

⁵The change in concept is not substantial. Indeed, we may regard $f_i(x, \mu)$ as a function of the $n + 1$ variables x, μ . In the x, μ space the domains of the barrier functions remain constant. Clearly, if the functions $f_i(x)$ are linear or convex quadratic, then so are the functions $f_i(x, \mu)$, and the properties of the logarithmic barrier function (the relative Lipschitz condition in [7] or self-concordance in [24]) also hold in the x, μ space. For other convex functions this is not always true. However, a complexity analysis based on these properties is not of our concern for the moment. We hope to identify larger classes of problems that can be solved by our method.

The particular perturbation we choose is

$$w := \rho g_0(x^0) + \sum_{i=1}^m \frac{1}{-f_i(x^0, 1)} g_i(x^0), \quad (3.4)$$

where ρ is a scalar to be specified later. (Both w and ρ are fixed throughout the algorithm.) For our barrier method we consider the functions $\varphi_\mu : P_\mu^o \rightarrow \mathbb{R}$,

$$\varphi_\mu(x) := \frac{\rho}{\mu} f_0(x) - \sum_{i=1}^m \log(-f_i(x, \mu)) - w^T x, \quad (3.5)$$

that combine a multiple of the objective function and the perturbed barrier function. For $\mu \in (0, 1]$ we define⁶ $x(\mu)$ to be a perturbed center if $x(\mu) \in P_\mu^o$ and if it is a minimum of the function φ_μ . We denote the gradient and Hessian of φ_μ by $g(x, \mu)$ and $H(x, \mu)$ as follows:

$$g(x, \mu) := \nabla \varphi_\mu(x) = \frac{\rho}{\mu} g_0(x) + \sum_{i=1}^m \frac{g_i(x)}{-f_i(x, \mu)} - w, \quad (3.6)$$

and

$$H(x, \mu) := \nabla^2 \varphi_\mu(x) = \frac{\rho}{\mu} H_0(x) + \sum_{i=1}^m \frac{H_i(x)}{-f_i(x, \mu)} + \frac{g_i(x)g_i(x)^T}{f_i(x, \mu)^2}. \quad (3.7)$$

Clearly, $H(x, \mu)$ is positive semidefinite if the f_i are convex. Here, we assume that it is positive definite. It becomes ill-conditioned, however, if x approaches some point on the boundary of P_μ at which less than n linearly independent constraints are active. In Section 4.4 we “expand” the matrix H and refer to [4] to improve the stability when solving linear systems with H . Note that $x = x(\mu)$ is a perturbed center if and only if it is a zero of the following characteristic equation:

$$g(x, \mu) = 0. \quad (3.8)$$

(Clearly, if x satisfies $g(x, \mu) = 0$, then by convexity it is a minimum of φ_μ . Conversely, if x is a minimum of φ_μ in the open set P_μ^o , then $g(x, \mu) = 0$.)

Note that by definition of w the point x^0 is the first center: $x^0 = x(1)$. Ideally we would like the perturbation w to be zero, in which case the points $x(\mu)$ are the analytic centers. The size of w depends on the choice of the starting point and on the shifts β_i . If w is close to zero, one can prove that the tangent to the curve $x(\mu)$ closely approximates the curve in some interval $[\mu_1, \mu_2]$ that does not depend on the problem data. For large w (measured in the norm given by $H(x, \mu)^{-1}$) this is no longer true. It is therefore important to initialize the method such that w is moderate in size, see Appendix A.2. Eventually, $H(x(\mu), \mu) \rightarrow \infty$ as $\mu \rightarrow 0$, so that w measured in the above norm tends to zero.

The function $\mu\varphi_\mu(x)$ is (at least for $\rho = 1$) just the objective function $f_0(x)$ to which a multiple (μ) of the barrier function is added, as in the outline of Section

⁶If P is empty then the definition is valid only for $\mu \in (\delta, 1]$, where δ is as in Lemma 1. We note that for certain degenerate cases the analytic center may not exist, while the perturbed center is well-defined.

2.1. Our choice of φ_μ in (3.5) rather than the seemingly more natural choice $\mu\varphi_\mu$ can be motivated by its gradient (3.6) defining the characteristic equation (3.8). If we set $\rho = 1$ for the moment, then (3.6) shows a certain symmetry in the treatment of g_0 and g_i for $i \geq 1$: The parameter μ may be considered as a function value $\mu = -f_0(x, \mu) := \lambda(\mu) - f_0(x)$ for some suitable quantity $\lambda(\mu)$. For quadratic f_i it has been shown in [29] that there is a monotone 1-1 correspondence between μ and $\lambda(\mu)$, and that $\lambda(\mu)$ converges to the optimal value $f_0(x^*)$. (The result generalizes in a straightforward way to convex f_i with self-concordant barrier functions.) Another reason why we consider φ_μ rather than $\mu\varphi_\mu$ is that the property of self-concordance is not invariant under multiplication of the barrier function by μ .

Under relatively weak assumptions the following properties can be shown.

Lemma 2

1. A unique perturbed center exists for all μ for which P_μ^o is not empty.
2. $\lim_{\mu \rightarrow 0} \min_{x \in P} \|x(\mu) - x\|_2 = 0$ if P is not empty.
3. $\lim_{\mu \rightarrow 0} f_0(x(\mu)) - f_0(x^*) = 0$ if x^* is an optimal solution to (CP) .

Proof: See Appendix C. ■

4. ONE ITERATION OF THE METHOD

The general idea of the method is as follows. Starting from $\mu = 1$ and $x(1) = x^0$, a sequence of iterates is generated in some neighborhood of the path $x(\mu)$. The iterates x^k are regarded as approximations to points $x(\mu^k)$ on the path of perturbed centers, where $\mu^0 = 1$, $\mu^k > 0$, $\mu^k \rightarrow 0$. The algorithm proceeds in three steps per iteration.

1. Compute the tangent x' to the curve $x(\mu)$ at the current iterate x^k and μ^k .
2. Choose adaptively a steplength $\alpha \in (0, 1)$ to follow the tangent starting from x^k . Let the resulting point be $\hat{x}^k = x^k - \alpha\mu^k x'$. Set $\mu^{k+1} = \mu^k(1 - \alpha)$. The steplength α is chosen such that $\hat{x}^k \in P_{\mu^{k+1}}^o$ and such that Newton iterations starting from \hat{x}^k for finding $x(\mu^{k+1})$ can be expected to converge rapidly.
3. Perform a small number of Newton steps to bring the iterate closer to the path of perturbed centers (using some "old" factorization of the Hessian and a linesearch with merit function $\varphi_{\mu^{k+1}}$). The result is x^{k+1} .

It is in Step 3 where our method differs from most implementations of interior-point algorithms for linear programming [4, 14, 15]. For linear programs it appears that the extra effort taken in Step 3 to move away from the boundary towards the center does not pay. For nonlinear problems our results indicate that "centering" stabilizes the algorithm and may be necessary in some cases.

4.1. The Tangent

Let $v(x, \mu)$ be the derivative of $g(x, \mu)$ with respect to μ :

$$v(x, \mu) := \frac{d}{d\mu} g(x, \mu) = -\frac{\rho}{\mu^2} g_0(x) - \sum_{i=1}^m \frac{\beta_i}{f_i(x, \mu)^2} g_i(x). \quad (4.1)$$

If $H(x(\mu), \mu)$ is positive definite the perturbed center is unique and the tangent to the curve of perturbed centers at $x(\mu)$ is defined by the linear system

$$H(x(\mu), \mu)x'(\mu) = -v(x(\mu), \mu). \quad (4.2)$$

Verification of (4.2) is straightforward by differentiating $g(x(\mu), \mu) \equiv 0$ in (3.6) with respect to μ . (It can be shown that $\lim_{\mu \rightarrow 0} \mu^2 H(x(\mu), \mu)$ exists and is nonzero. In our implementation we therefore use $\mu^2 H(x, \mu)$ and $\mu^2 v(x, \mu)$ instead of the unbounded quantities $H(x, \mu)$ and $v(x, \mu)$.)

Note that w does not occur in the definition of the tangent. If the current iterate is some point x^k that is not on the path of perturbed centers, then the above quantity is the tangent to some other perturbed center curve that also leads to an optimal solution x^* .

We note the inherent sparsity of $H(x, \mu)$ if the functions f_i each depend on few variables only, or if there is a small number of separable functions f_i . (For separable f_i the gradient g_i and thus also $g_i g_i^T$ could be full. The "expansion" of H suggested in Section 4.4 that is intended to reduce the condition number of the linear system also preserves the sparsity in this case.)

In Step 1 above we determine x' from (4.2) with x^k in place of $x(\mu)$ and $\mu = \mu^k$, i.e. from the system $H(x^k, \mu^k)x' = -v(x^k, \mu^k)$. The steplength α in Step 2 depends on how well Newton's method converges. We focus on the Newton step first.

4.2. The Newton Step

The Newton step Δx for finding $x(\mu)$ starting from $x \in P_\mu$ is given by the system

$$H(x, \mu)\Delta x = -g(x, \mu). \quad (4.3)$$

From [24, 6] we know that Newton's method (without linesearch) for finding the center $x(\mu)$ converges quadratically if φ_μ is self-concordant (i.e. if the Hessians of the constraints satisfy a relative Lipschitz condition) and

$$\gamma := \Delta x^T H(x, \mu) \Delta x = g(x, \mu)^T H(x, \mu)^{-1} g(x, \mu) = -g(x, \mu)^T \Delta x < \frac{1}{4}. \quad (4.4)$$

We note that a Newton step for finding $x(\mu^{k+1})$ may not be necessary, since as mentioned above, all the "perturbed center curves" end in the optimal set, and one could continue by following the tangents of different curves. However, the step along the tangent may bring the point \hat{x}^k close to the boundary of P_{μ^k} (and iterating too close to the boundary of P_μ slows down convergence), so that a Newton correction is indeed useful. In our program we compute the Newton correction with the same factorization of H as already used for the computation of x' . The linesearch during the Newton step is controlled by the merit function $\varphi_{\mu^{k+1}}(x)$.

4.3. The Steplength α During Extrapolation

Our goal is to choose $\alpha \in (0,1)$ as large as possible such that the extrapolation $\hat{x}^k = x^k - \alpha \mu^k x'$ is strictly feasible, i.e. $\hat{x}^k \in P_{\mu^{k+1}}^o$ with $\mu^{k+1} = \mu^k(1 - \alpha)$, and such that Newton's method for finding the (next) center converges rapidly.

An obvious possibility to guarantee the second condition is to choose the step-length during extrapolation small enough such that the first Newton step Δx starting at the predicted point satisfies a relation of the type (4.4). However, this results in very short steps α in general. In our implementation we first approximated the maximum possible steplength α_{\max} such that $x + \alpha_{\max} x'$ is still feasible and then took r percent of α_{\max} . If it turned out that Newton iteration converged quickly we increased r for the next extrapolation, and conversely, if Newton iteration was slow we decreased r .

4.4. Improving the Stability of the Linear Systems

At the definition of H in (3.7) we mentioned the instability to be expected when solving systems with H . In [4], Gill et al. present an approach to stabilize the solution of KKT systems. To illustrate how their analysis applies to our matrix H we set $\rho = 1$ and $\beta = 0$ for the moment. Note that H can be written in the form

$$H(x, \mu) = \hat{H}(x, \mu) + J^T D^{-2} J. \quad (4.5)$$

where $\hat{H}(x, \mu) = \frac{1}{\mu} H_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} H_i(x)$, J is the Jacobian of the constraints,

$$J = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}, \quad (4.6)$$

and $D = \text{diag}(f_i(x))$. Solving a system with H is equivalent to solving a system with

$$K = \begin{pmatrix} -D^2 & J \\ J^T & \hat{H} \end{pmatrix}, \quad (4.7)$$

which can be seen when taking the Schur complement of $-D^2$ within K . This system in turn is equivalent to a system involving

$$K' = \begin{pmatrix} D^{-2} & 0 & I \\ 0 & \hat{H} & J^T \\ I & J & 0 \end{pmatrix}. \quad (4.8)$$

(Take the Schur complement of D^{-2} .) Systems of the form K' are considered in [4]. The basic idea is that it is better to factorize K' directly, or to take the Schur complement of just certain parts of the diagonal matrices, such that the Schur complement does not become excessively ill-conditioned (and does not suffer excessive fill-in caused by dense rows or columns of J).

4.5. Stopping test

Let ϵ be the desired final accuracy in the objective function. A possible stopping criterion is $\mu \leq \epsilon \rho(1 + \|Df_0(x)\|)/m$. For convex constraints with self-concordant barrier functions and a linear objective function this stopping test is exact at points on the path of analytic centers [7]. However, since the constraints have been shifted, the final iterate is not always feasible. It is only guaranteed that $f_i(x) \leq \mu\beta_i$. The relative constraint violation is bounded by $\mu\beta_i/(1 + \|Df_i(x)\|)$. In the numerical experiments below we have chosen

$$\bar{\mu} := \min(\epsilon \rho(1 + \|Df_0(x^0)\|)/n, \epsilon(1 + \|Df_i(x^0)\|)/(\beta_i + \epsilon))$$

for $1 \leq i \leq m$ and stopped as soon as $\mu \leq \bar{\mu}$.

4.6. Convergence

Before concluding this description we briefly state some convergence results.

- If (CP) has an optimal solution, one can show under weak conditions that as $\mu \rightarrow 0$ the iterates x^k satisfy the same limit relations (for $k \rightarrow \infty$) as stated for $x(\mu)$ in Lemma 2.
- If (CP) has no optimal solution, either $x^k \rightarrow \infty$ (if the problem is unbounded) or we find that $\mu \rightarrow \delta > 0$ with δ as in Lemma 1 Part 4. Both cases ($x^k \rightarrow \infty$ and $\mu \rightarrow \delta > 0$) are hard to identify in an implementation and need special attention.

5. NUMERICAL EXPERIMENTS

The above method was implemented in MATLABTM[19] and tested on a few problems with up to 300 unknowns and dense arithmetic. As mentioned before, the use of sparse-matrix techniques will be crucial for the efficiency of this method. The development of efficient interior-point methods for linear programs took several years and similar efforts may be needed for developing an interior-point method for non-linearly constrained problems. The goal of the implementation here was merely to illustrate the behavior of the method in terms of number of iterations and Newton corrections, and to test various parameters (such as β and ρ) that define the barrier function.

The statistics below read as follows. Each iteration involves computation of the tangent and a small number (1 to 10) of Newton steps. The tangent and the Newton steps are computed from a linear system that involves the Hessian of φ . Sometimes more than one Hessian is needed in an iteration. Each Hessian is used for several Newton steps; their computation and factorization dominates the overall computation.

5.1. Problem Manne

This Problem is taken from [21], where it is presented in two versions. The first version involves 300 variables, a logarithmic objective function, 100 nonlinear constraints, 100 linear inequalities and 400 simple bounds. The second version is identical except that it has only 300 simple bounds.

In [21] results are given for MINOS. Version one took 7 major iterations, 183 minor iterations, 497 function evaluations and 12 seconds on an IBM 370/168, while version two required 11 major iterations, 355 minor iterations, 859 function evaluations and 34 seconds.

MINOS performs best if a high number of linear constraints or bounds are active in the optimal solution, thus reducing the size of the (dense) systems that are solved in each iteration. For version one, the size of the dense systems grew to 25, and for version two they grew to 99 (since some of the active bounds in version one were removed).

In contrast to MINOS, the size of the systems to be solved in each iteration of the interior-point algorithm is always 300, i.e. the number of variables. For both versions of the problem these systems are sparse and of diagonal structure, with at most 7 nonzeros per row. A sparse-matrix solver could take great advantage of the structure.

Problem Manne is sparse and convex, but it does not satisfy the conditions of [6] or [24] that guarantee a fixed minimum rate of convergence.

Below we report the results of our method for version one and version two. As starting point we chose the (infeasible) vector of all ones. (The objective and some of the constraints are not defined for $x = 0$.) In contrast to MINOS, the interior-point algorithm performed slightly better for the second problem, giving hope that for certain problems in which the active constraints do not significantly reduce the dimension of the MINOS subproblems, interior-point algorithms may become an attractive alternative.

Problem Manne was one for which our initialization in Section 3.3 resulted in a vector w of norm $(w^T H^{-1} w)^{1/2} \approx 15$. By the procedure in Appendix A.2 we decreased the norm of w to about one before starting the iterations. For comparison we also list the results for problem one without reducing the size of w . In this case convergence was very slow, and for many iterations the maximum steplength α_{\max} during extrapolation was less than 0.25. In all examples, Newton's method for finding the perturbed center at each iteration was terminated when the H -norm $\|\Delta x\|_H := (\Delta x^T H(x, \mu) \Delta x)^{1/2}$ of the Newton step Δx was less than $1/2$.

We also show the results for a smaller version of problem one with only 30 unknowns, to show that the number of iterations grows only moderately with the number of variables for this particular problem.

Problem	Large w	Version one	Version two	Small problem
Iterations	38	19	15	11
Hessians	40	24	20	13
Newton steps	115	76	61	38
$\ \beta\ _\infty$	2.05	2.05	2.05	2.05
ρ	4.1	4.1	3.8	0.29
Final μ	3.4e-9	5.7e-9	2.2e-8	8.7e-9
Objective	9.287556	9.287556	9.330183	2.670098
Constraint violation	4.9e-9	8.1e-9	3.1e-8	1.1e-8

5.2. Problem 385, Schittkowski

This is a problem with 15 unknowns, 10 convex quadratic constraints and a linear objective function. It is taken from [27], where it was solved with NLPQL [26] using 693 function evaluations and 242 gradient evaluations. Running times or numbers of arithmetic operations are not reported in [27]. The starting point (zero) was strictly feasible and β was zero. (Hence, P_μ was constant and also the final point was strictly feasible.) Our implementation took 11 iterations to solve the problem, a total of 11 evaluations of the Hessian, 27 Newton steps (each of which requires the evaluation of the gradient of φ) and 64 additional gradient evaluations for the linesearch steps. The steplength α was 0.80 on average, ranging from 0.70 to 0.92. The Hessians of the constraints are diagonal, but to preserve the sparsity of the Hessians of φ , the dense outer products of the gradients in (3.7) must be treated separately (for example as in Section 4.4).

5.3. Problem 386, Schittkowski

This is the same as problem 385 above (except that two entries in the coefficients of the constraints are changed) with an additional *concave* constraint. In [27], 900 function evaluations and 327 gradient evaluations were required to solve the problem to 6 digits of accuracy. In order to explore the limit of applicability of our method we tested this problem with different parameter settings.

- Using the standard method, the Hessian of φ became indefinite in the 8-th iteration and our algorithm failed.
- In a second run we set the Hessian of the *concave* constraint equal zero but kept all other second-derivative information. The method converged to the true solution in 10 iterations using 10 evaluations of the Hessian of φ and 28 Newton steps.
- In a third run we set the Hessians of all constraints to zero (simulating a linearized problem). In this case, the Hessian of φ was indefinite to begin with (since there were only 11 linearized constraints in a 15 dimensional space) and the method failed again.
- Finally we replaced the Hessians of all constraints by multiples of the identity (such that the norm of the replacement approximately equals the norm of the

true Hessian). This time the method took 15 iterations, 20 evaluations of the Hessian of φ and 62 Newton steps.

5.4. Conclusions

The design of fast and stable implementations of interior-point algorithms is marked by a number of conflicting principles.

- It is desirable to maintain some polynomiality results that limit the dependence of the method on the data of a particular problem. However, most polynomial-time interior-point methods are too conservative in the choice of the steplength and therefore inefficient in practice.
- To maintain affine invariance of the method seems equally important, to reduce the influence of affine transformations of the problem. With finite-precision arithmetic this influence however cannot be eliminated completely.
- Finally, the linear systems involved should be kept well-conditioned.

A typical example of how to take these concepts into account in the above method is the steplength along the tangent: the closer the extrapolation to the boundary, the more ill-conditioned the Hessian of the barrier function (and the worse the theoretical complexity), suggesting that one should not take too large steplengths. The concept of numerical stability based on the condition number of a matrix however is not perfect. Not only do interior-point methods for some reason perform well when taking steps of 99.995% to the boundary (!!!) [14], but there are also simple examples for which the condition number of a matrix is (almost) irrelevant; for example when solving equations with the “ill-conditioned” diagonal matrix $\text{diag}(10^{10}, 10^{-10})$.

A. DETAILS FOR THE INITIALIZATION

A.1. Additional Assumption

In the footnote preceding the definition of the functions $f_i(x, \mu)$ (3.1) we have referred to Appendix A for a more precise statement about our assumptions. To guarantee convexity we require the following assumption.

We assume that the functions f_i are convex and continuous in the set P_1 and smooth in its interior. That is, we assume that $P_1 \subset S$.

If this assumption is violated we can try to change the given values of β_i or the starting point x^0 . (P_1 depends on β_i and x^0 .)

A slightly more complicated modification is as follows. If points \bar{x}^i are known such that $f_i(\bar{x}^i) < 0$, we may consider the functions $f_i(x, \mu) := f_i(x + \mu(\bar{x}^i - x^0))$. In this case, Lemma 1, Part 1 and 2 no longer hold but the sets P_μ still converge to P and Part 3 and 4 still hold. We will not discuss this modification any further but concentrate on the hopefully more common case (3.1).

A.2. Decreasing the Perturbation

As mentioned in Section 3.3 it is important that the perturbation w describing the perturbed centers not be too large; more precisely that $w^T D^2 \phi_1(x^0)^{-1} w$ is say less than 1. Only for small vectors w is it possible to prove polynomiality for linear constraints, and our numerical experiments suggest that in practice also, large perturbations w slow down convergence. In some examples however, the initialization outlined in this paper does yield large perturbations w , and it is necessary to reduce the size of w before starting the algorithm. For such cases, our implementation reduced the size of w by the following procedure.

- Before starting the predictor-corrector iterations set $w = 0$ and introduce additional constraints of the form $(x_i - x_i^0)^2 \leq 10^{12} t^2$ (with t as in Section 3.1).
- Perform a number of Newton steps for finding the (analytic) center of P_1 with the additional constraints, and stop when the H -norm $\|\Delta x\|_H = (\Delta x^T H(x, 1) \Delta x)^{1/2}$ of the Newton step Δx satisfies a given bound. Let the result be \bar{x}^0 .
- Remove the additional constraints again and redefine w for the new starting point \bar{x}^0 as outlined in Section 3.3.

This procedure does not assume that a bound of the form " $\|x - x^0\|_\infty \leq 10^6 t$ for all feasible x " is known a priori; the additional constraints are only used for decreasing w to guarantee that Newton's method is well defined, and they are later removed.

A.3. Warm Start

If an initial point x^0 is given that is "almost" optimal, a "warm start" is possible by defining the quantity t (before (3.1)) as $\max\{10^{-4}, f_i(x^0)\}$ for example, rather than $\max\{1, f_i(x^0)\}$, and by fixing $\rho \geq 1$ to minimize the norm of the gradient of $\varphi_1(x^0)$.

We point out a possible problem that may occur with this warm start. If the initial point x^0 is feasible, then the sets P and P_μ coincide—up to a perturbation of the size 10^{-4} . In contrast to many interior-point methods for linear programs, our method requires strict fulfillment throughout of all inequalities describing the set P_μ (since outside P_μ the functions f_i may not be defined or may not be convex). If the set P is "long and thin" (e.g. a constraint of the form $|x_1| \leq 10^{-5}$ that bounds the first component of the vector x makes the set P very thin) and the initial point x^0 is feasible but far from the optimum, then it may be wise to define β_i as outlined in (3.1), rather than using a warm start and "winding in a long thin neighborhood of the path of centers" in P from x^0 to the optimal solution⁷. The definition in (3.1) is chosen to avoid such a long thin path.

⁷A convex set P being long and thin is no problem in theory; there is always an affine mapping that maps P into a "nearly round set" (i.e. into a set that contains a ball of radius 1 and is contained in a ball of radius n , the dimension of the space). However, the affine mapping changes the condition numbers of the Hessians. Hence, in the context of finite precision, the concept of affine invariance is only of limited relevance.

A.4. The Multiplier ρ

In the definition of the center (3.5) we did not elaborate on the choice of the multiplier ρ for the objective function. However, a good choice of ρ is very important. For one-dimensional examples it is easy to see that a poor choice can result in the curve of centers passing the same point x twice (which is not attractive when following this curve numerically). We applied the following heuristic in our implementation.

- If the constraints have been shifted, i.e. if $\beta \neq 0$, we compute a tangent t_1 for the case $\rho = 0$ and a tangent t_2 for the case that $\rho = 1$ and that the shift is kept unchanged throughout the method. We then choose $\rho = \|t_2\|/\|t_1\|$.
- If the constraints have not been shifted, i.e. if $\beta = 0$, then we choose ρ such that the H -norm $\|x'\|_H$ of the tangent x' at the first iteration is 1.

B. SOME RESULTS ON THE SETS P_μ

Proof of Lemma 1:

1. Clear, because $\beta_i \geq 0$.
2. For $x \in P_1 \setminus P$ $\exists i : f_i(x) > 0$. Hence $f_i(x) - \mu\beta_i > 0$ for small $\mu > 0$; i.e. $x \notin P_\mu$ for such μ .
3. Let $x^1 \in P$; i.e. $f_i(x^1) \leq 0$, $1 \leq i \leq m$. Then $x^1 + \mu(x^0 - x^1) \in P_\mu^o$ because

$$f_i(x^1 + \mu(x^0 - x^1), \mu) = f_i(\mu x^0 + (1 - \mu)x^1) - \mu\beta_i$$

$$\leq \mu f_i(x^0) + (1 - \mu)f_i(x^1) - \mu(f_i(x^0) + t) \leq -\mu t < 0.$$
4. Let $\delta := \inf_{\mu \geq 0} P_\mu \neq \emptyset$. By continuity of f_i , $\delta < 1$. If $\mu > \delta$, there is a $\bar{\mu} \in [\delta, \mu)$ for which $P_{\bar{\mu}}$ is not empty. Similarly to Part 3, it then follows with $x^1 \in P_{\bar{\mu}}$ that $P_\mu^o \neq \emptyset$. (We note that $\delta = 0$ is possible if P_1 is unbounded.) ■

Part 2 of Lemma 1 also holds in the following stronger form.

Lemma 3

Let $K_r := \{x \mid \|x\|_2 \leq r\}$ and $\mathbf{P}_\epsilon := \{x \mid x = y + z, y \in P, \|z\|_2 \leq \epsilon\}$. Then for all finite r and all positive ϵ there is a positive μ such that

$$P_\mu \cap K_r \subset \mathbf{P}_\epsilon,$$

which implies that P_μ converges uniformly to P in any ball K_r . The restriction to a bounded set K_r is necessary, since there are examples for which $P_\mu \not\subset \mathbf{P}_1$ for any $\mu > 0$.

Proof: We prove the first statement by contradiction. Suppose there was a finite r and a positive ϵ such that for all $\mu > 0$, $P_\mu \cap K_r \not\subset \mathbf{P}_\epsilon$. Let $\mu^k \rightarrow 0$, $\mu^k \in (0, 1)$ be a sequence and $x^k \in (P_{\mu^k} \cap K_r) \setminus \mathbf{P}_\epsilon$. Since $\|x^k\| \leq r$ there is an accumulation point \bar{x} . Clearly $\bar{x} \in P$ (otherwise $\exists i_0 : f_{i_0}(x) > 0$ and then by continuity of f_{i_0} ,

$\exists \delta, \sigma > 0 : f_{i_0}(\bar{x} + K_\delta) > \sigma$, contradicting the definition of \bar{x} . By construction we also know that $\bar{x} \notin P_c^\circ$, in contradiction to $P \subset P_c^\circ$. ■

That the more general statement " $P_\mu \subset P_c$ for small enough μ " is not true can be seen from a simple counterexample. Take the function $f(x, y) := y^2/x$ with domain $S := \{(x, y) | x \geq 1\}$, defining $P := \{(x, y) \in S | y^2/x \leq 0\}$. It is easy to verify that f is convex and that $P_\mu = \{(x, y) | x \geq 1, |y| \leq \sqrt{\mu x}\} \not\subset P_1$ for any $\mu > 0$.

Note that $\min\{|y| | (x, y) \in P\}$ exists in this case, but not so $\min\{|y| | (x, y) \in P_\mu\}$ for $\mu > 0$, and neither does a perturbed center exist for $0 < \mu < 1$. (This example shows another surprising property. If f_1 and f_2 are convex functions that each have a minimum on a common closed set S , then $f_1 + f_2$ may not have a minimum on S ; e.g. take y^2/x and $(y-1)^2/x$ on the above set S .) ■

Despite this counterexample, the vague intuition that P_1 might not be much "bigger" than P can be formalized in the following simple lemma.

Lemma 4

If P_δ is nonempty and bounded for some $\delta \in [0, 1)$, then so is P_1 .

Proof: (by contradiction.) Without loss of generality let $\delta = 0$ and $x^1 = 0 \in P$. (Then also $0 \in P_1$.) Suppose now that P_1 is unbounded. By convexity of P_1 there is a vector s such that $\lambda s \in P_1$ for all $\lambda \geq 0$. Boundedness of $P = P_0$ implies that there exists $M < \infty$ such that $\lambda s \notin P$ if $\lambda \geq M$, i.e. $\exists i_0 : f_{i_0}(Ms) > 0$. We conclude that $\delta := (f_{i_0}(Ms) - f_{i_0}(0))/M > 0$, and hence for $\lambda \geq M$ it follows from convexity of f_{i_0} that

$$f_{i_0}(\lambda s) \geq f_{i_0}(Ms) + (\lambda - M)\delta \rightarrow \infty$$

as $\lambda \rightarrow \infty$. This implies that $f_{i_0}(\lambda s) - \beta_i > 0$ for sufficiently large λ , which in turn implies that $\lambda s \notin P_1$ for such λ , in contradiction to the choice of s . ■

C. CONVERGENCE OF THE CENTERS

We are now ready to show briefly the results stated in Lemma 2, at least for certain cases. The results hold under more general conditions, but the most general conditions are not of concern in the present paper.

Proof of Lemma 2

Throughout this lemma we assume that P_1 is bounded (which is the case, for example, if P is nonempty and bounded), and that φ_1 is strictly convex (as for linear or quadratic or self-concordant programs).

First we note that a perturbed center $x(\mu)$ exists for all μ for which P_μ° is not empty, since $-\sum_{i=1}^m \ln(-f_i(x, \mu)) \rightarrow \infty$ as x approaches the boundary of P_μ . If φ_1 is strictly convex then so is φ_μ (as long as $P_\mu^\circ \neq \emptyset$), and the perturbed center is unique.

The second part of Lemma 2 follows from Lemma 3 since $x(\mu) \in P_\mu \subset P_1$ and P_1 is bounded, i.e. $P_1 \subset K_r$ for some r .

To prove the third part we note that when multiplying (3.5) by μ we obtain

$$x(\mu) = \arg \min_{x \in P_\mu} \mu \varphi_\mu(x) = \arg \min_{x \in P_\mu} f_0(x) - \mu \sum_{i=1}^m \ln(-f_i(x, \mu)) - \mu w^T x.$$

Let $x^\mu := x^* + \mu(x^0 - x^*)$. We conclude as in the proof of Lemma 1 Part 3 that $f_i(x^\mu, \mu) \leq -\mu t$ for all i . This implies that

$$-\mu \sum_{i=1}^m \ln(-f_i(x^\mu, \mu)) - \mu w^T x^\mu \leq -m\mu \ln(\mu t) - \mu w^T x^\mu \rightarrow 0$$

as $\mu \rightarrow 0$. (Note that $w^T x^\mu$ is bounded for $\mu \in [0, 1]$.) By continuity of f_0 it also follows that $f_0(x^\mu) \rightarrow f_0(x^*)$, so that $\limsup_{\mu \rightarrow 0} \mu \varphi_\mu(x^\mu) \leq f_0(x^*)$. From Part 2 above and continuity of f_0 it follows that $\liminf_{\mu \rightarrow 0} f_0(x(\mu)) \geq f_0(x^*)$. Further, $\mu \varphi_\mu(x(\mu)) \leq \mu \varphi_\mu(x^\mu)$, and since the logarithmic barrier terms $-\ln(-f_i(x(\mu), \mu))$ and $w^T x(\mu)$ are bounded below for $\mu \in [0, 1]$, the claim follows from the above inequalities. ■

References

- [1] A. Fiacco, G.P. McCormick, "*Sequential unconstrained minimization techniques*", Nonlinear Programming, (Wiley, New York, 1968).
- [2] K.R. Frisch "The logarithmic potential method of convex programming", Technical Report, University Institute of Economics (Oslo, Norway, 1955).
- [3] P.E. Gill, Private communication.
- [4] P.E. Gill, W. Murray, D.B. Ponceleón, M.A. Saunders, "Solving reduced KKT systems in barrier methods for linear and quadratic programming", Report SOL 91-7, Dept. of Operations Research, Stanford University, Stanford, CA (1991).
- [5] F. Jarre, "On the convergence of the method of analytic centers when applied to convex quadratic programs", Report No. 35 (1987), Schwerpunktprogramm der DFG Anwendungsbezogene Optimierung und Steuerung. To appear in *Mathematical Programming* 49 (1991).
- [6] F. Jarre, "The method of analytic centers for solving smooth convex programs", in S. Dolecki (ed.), *Lecture Notes in Mathematics*, Vol. 1405, Optimization, Springer (1989) 69-85.
- [7] F. Jarre, "Interior-point methods for convex programming", Report SOL 90-16, Dept. of Operations Research, Stanford University, Stanford, CA (1990).
- [8] F. Jarre, M.A. Saunders, "An adaptive interior-point method for linear programming", Report SOL 91-1, Dept. of Operations Research, Stanford University, Stanford, CA (1991).
- [9] F. Jarre, G. Sonnevend, J. Stoer, "An implementation of the method of analytic centers", in A. Benoussan, J.L. Lions (eds.), *Lecture Notes in Control and Information Sciences* 111, Springer, New York (1988).
- [10] N. Karmarkar, "A new polynomial-time algorithm for linear programming", *Combinatorica* 4 (1984) 373-395.
- [11] M. Kojima, N. Megiddo, T. Noma, A. Yoshise, "A unified approach to interior-point algorithms for linear complementarity problems", Report RJ 7493 (70008), IBM Research, Almaden Research Center, San Jose, CA (1990).
- [12] M. Kojima, S. Mizuno, T. Noma, "A new continuation method for complementarity problems with uniform P-functions," Report B-194, Dept. of Information Sciences, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, (1987).

- [13] M. Kojima, S. Mizuno, A. Yoshise, "A little theorem on the big M in interior-point algorithms," Report B-239, Dept. of Information Sciences, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, (1991).
- [14] I.J. Lustig, R.E. Marsten, D.F. Shanno, "On implementing Mehrotra's predictor-corrector interior-point method for linear programming", Report SOR 90-03, Dept. of Civil Eng. and OR, Princeton University, Princeton, NJ (1990).
- [15] S. Mehrotra, "On the implementation of a (primal-dual) interior-point method", Report 90-03, Dept. of Ind. Engineering and Management Sciences, Northwestern University, Evanston, IL (1990).
- [16] S. Mehrotra, J. Sun, "An interior-point algorithm for solving smooth convex programs based on Newton's method", Report 88-08, Dept. of Ind. Engineering and Management Sciences, Northwestern University, Evanston, IL (1988).
- [17] N. Megiddo, "Pathways to the optimal set in linear programming", in N. Megiddo (ed.), *Progress in Mathematical Programming: Interior-Point and Related Methods*, Springer Verlag, New York (1989) 131-158.
- [18] J. Mennicken, "Implementation of a first order central path following algorithm for solving large linear programs", Report No. 202, Schwerpunktprogramm der DFG Anwendungsbezogene Optimierung und Steuerung, Institut für Ang. Math. und Statistik, Universität Würzburg, Am Hubland (1990).
- [19] C. Moler, J. Little, S. Bangert, "PRO-MATLAB user's guide", The Math-Works, Inc., Sherborn, MA (1987).
- [20] W. Murray, "Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions", *Journal Opt. Theory Appl.*, vol. 7, 189-196 (1971).
- [21] B.A. Murtagh, M.A. Saunders, "A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints", *Mathematical Programming Study* 16 (1982) 84-117.
- [22] B.A. Murtagh, M.A. Saunders, "MINOS 5.1 User's guide", Report SOL 83-20R, Dept. of Operations Research, Stanford University, Stanford, CA (1983, 1987).
- [23] J.E. Nesterov, A.S. Nemirovsky, "A general approach to polynomial-time algorithms design for convex programming", Report, Central Economical and Mathematical Institute, USSR Acad. Sci., Moscow, USSR (1988).
- [24] J.E. Nesterov, A.S. Nemirovsky, "Self-concordant functions and polynomial-time methods in convex programming", Report, Central Economical and Mathematical Institute, USSR Acad. Sci., Moscow, USSR (1989).
- [25] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming", *Mathematical Programming* 40 (1988) 59-93.
- [26] K. Schittkowski, "NLPQL: A FORTRAN subroutine solving constrained nonlinear programming problems", *Annals of Operations Research* 5, 485-500 (1985).
- [27] K. Schittkowski, "More Test Examples for Nonlinear Programming Codes", M. Beckmann and W. Krelle (eds.), *Lecture Notes in Economics and Mathematical Systems* 282, Springer (1987).
- [28] G. Sonnevend, "An 'analytical centre' for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming", *Lecture Notes in Control and Information Sciences* 84 (1986) 866-878.
- [29] G. Sonnevend, J. Stoer, "Global ellipsoidal approximations and homotopy methods for solving convex analytic programs", *Appl. Math. and Optimization* 21 (1989) 139-166.
- [30] M.H. Wright, "Numerical methods for nonlinearly constrained optimization", Ph.D. Thesis, Computer Science Dept., Stanford University, Stanford, CA (1976).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE July 1991		3. REPORT TYPE AND DATES COVERED Technical Report	
4. TITLE AND SUBTITLE Practical Aspects of an Interior-Point Method for Convex Programming				5. FUNDING NUMBERS N00014-89-J-1659	
6. AUTHOR(S) Florian Jarre and Michael A. Saunders					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Operations Research - SOL Stanford University Stanford, CA 94305-4022				8. PERFORMING ORGANIZATION REPORT NUMBER 1111MA	
9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research - Department of the Navy 800 N. Quincy Street Arlington, VA 22217				10. SPONSORING / MONITORING AGENCY REPORT NUMBER SOL 91-9	
12a. DISTRIBUTION AVAILABILITY STATEMENT UNLIMITED				12b. DISTRIBUTION CODE UL	
13. ABSTRACT (Maximum 200 words) We present an algorithm for solving convex programs with nonlinear constraints. The algorithm works in the primal space only and uses a predictor-corrector strategy to follow a smooth path that leads to an optimal solution. The algorithm simultaneously iterates towards feasibility and optimality. The matrices occurring in the algorithm can be kept sparse if the nonlinear functions are separable or depend on few variables only. Some promising numerical examples obtained from a preliminary implementation are included.					
14. SUBJECT TERMS convex program; interior-point method; implementation; sparsity				15. NUMBER OF PAGES 20 pps.	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT SAR		